MAKING FABLABS SMART THROUGH SENSORS AND BIG DATA ANALYSIS

Dipl.-Ing. Friessnig M., BSc; Dipl.-Ing. Böhm T., BA; Univ.-Prof. Dipl.-Ing. Dr. Ramsauer C.
Institute of Industrial Management and Innovation Research, Graz University of Technology, Austria

matthias.friessnig@tugraz.at; thomas.boehm@tugraz.at

Abstract: *FabLabs are makerspaces, where you can transform your sketch into a real product with easy-to-use machines such as 3D printers, laser cutters, or desktop CNC mills. However, FabLabs, like most production environments, have a problem: the digital manufacturing machines provide only a one-way-interaction: human to computer to machine – from bits to atoms. A 3D printer, for example, produces objects unaware of its environment. But, what if we could change the machine so it could react to outside influences, such as other machines, people, or environmental conditions? This is the starting premise for the project in which we implemented a vast number of sensors at FabLab Graz and carried out Big Data Analyses. The research project is structured along the Big Data approach under supervision of Prof. Viktor Mayer-Schönberger from Oxford University, and is divided into three main project phases: First, integration of sensors in the laser cutter, 3D printer, CNC mill and also the environment of FabLab Graz. Second, the collection, analysis and processing of the collected data. In the third phase we focus on the derivation of use cases and product improvements based on data correlation. The aim of the project is to establish a smart FabLab that provides insights into operational enhancements and new product concepts to optimize machine interaction. These product concepts and the basic set up of the sensor system are open source and can be implemented easily at other FabLabs throughout the worldwide community. The paper describes the first and second phase of this ongoing research project in detail and gives first insights in phase three. Data, however, is unstoppable and while it is being gathered new correlations or enhancements can be discovered even in the course of writing this sentence.*

Keywords: RAPID PROTOTYPING, 3D-PRINTING, 3D-SCANNING, GENERATIVE METHODS, STEREOLITHOGRAPHY, GENERATIVE SINTERING, FUSED DEPOSITION MODELING, FABRICATION LABORATORY, PRODUCT INNOVATION PROJECT, FAB LAB MOVEMENT

1. Introduction

The FabLab movement is currently spreading rapidly around the world and there are already more than 540 registered FabLabs. [1] The second FabLab in Austria was established at the Institute of Institute of Industrial Management and Innovation Research at Graz University of Technology in 2014. The equipment was well chosen following the Massachusetts Institute of Technology guidelines. The core FabLab Graz equipment includes a laser cutter (Universal VLS 3.50), a plotter cutter (Roland CAMM-1 Servo GX-24), a computer numerical control (CNC) mill (Roland MDX-540) and two 3D printers. One uses the Fused Deposition Modeling (3D Touch) technology, and the other one uses the stereolithography method (Form 1+). [2]

As a part of the opening of FabLab Graz, a team was formed to start a research project concerning Big Data Analysis in connection with digital fabrication machines. The goal was to make FabLab machinery smarter and more efficient. One of the most popular machines in FabLabs is a 3D printer. 3D printing has gained enormous and increasing popularity since 2003 due to the declining cost of operation. It is used in various fields such as jewelry, education, architecture, industrial design etc. [3] To increase the performance of 3D printer and other digital manufacturing machines, the project team focuses on Big Data Analytics and the implementation of a sensor system. “Big Data is not only about storing data, but it also provides powerful real-time analytics and visualization tools”. [4] The crust of this approach is that future industry will encompass “smart machines, storage systems and production facilities capable of autonomously exchanging information, triggering actions and controlling each other independently”. [5] The Big Data approach is not a new one; in fact, many leading companies worldwide immensely dependent on Big Data analyses in their daily operations such as Google, Amazon, Wal-Mart and even the US government. [6] According to a study conducted by Strategy& PwC in 2014, more than 80% of companies will have digitized their value chain in the next five years. The study shows that industrial Internet enhances productivity and resource efficiency, with 18% projected increase in efficiency within five years. [7]

The novelty of this research project lies in the fact that while this approach is being implemented on a vast scale by industry, it still remains unpopular among FabLab communities. We prove that there is a significant potential to improve FabLab’s utilization by connecting FabLab machines with their environment. So FabLabs become smarter and more interactive.

This paper presents the main findings of this ongoing research project from the period of October 2014 until June 2015. The main project task in this period was the integration of sensors in the machines of FabLab Graz to collect, analyze and process data, and gain first insights that improve FabLabs, as a part of a feasibility study for further research.

2. Big Data Approach

This section provides the theoretical background for the Big Data analytics approach of the project. Big Data is defined as volumes of data available in varying degrees of complexity, generated at different intervals of time and varying degrees of ambiguity, which can scarcely be processed at all using traditional technologies, processing methods, algorithms, or any commercial off-the-shelf solutions. [8] In this context, volume names the ever growing size of data, variety highlights the growing number of data sources, and intervals of time reflect the speed in which data is generated. These three attributes build the foundation of a technological definition of Big Data. [9] [10]

In contrast to this academic definition for Big Data, Microsoft gives a practical example for it “a single customer for the company’s customer support monitoring and analysis operations can have thousands of the company’s sensors and programmable logic controllers, each delivering up to tens of thousands data points per second measuring temperature, pressure, vibration, etc. That’s big data.” [11] Based on this framework and on the ELTA (Extract, Load, Transform and Analyze) model by Marín-Ortega et al. the Big Data approach of this research project is shown in Figure 1. [12]

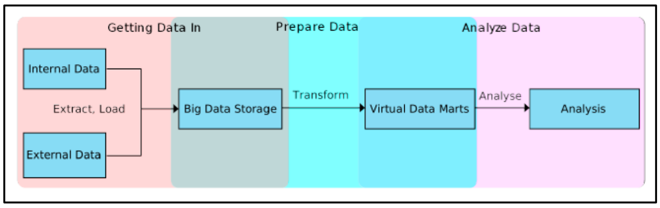

Figure 1: Extract, Load, Transform and Analyse (ELTA) [12]

This framework illustrates the data flow within an organization and shows the tasks performed on data. Data is first extracted based on the information needs and loaded into a Big Data store. It is then transformed out of the Big Data store and saved in a Virtual Data Mart (for fast access and processing). Finally, data from the Virtual Data Mart is used to perform analytic actions. [12]

Bringing the Data In (Extract and Load)

Data can be divided into structured and unstructured data. Data is structured if the format is known up front and no changes to this format ever happen. On the other hand, unstructured data such as texts do not follow any predetermined structures. This type of data occupies enormous storage volumes in every organization. Because of its large volume, unstructured data is seen as Big Data in an internal context. [13]

Data can be further divided into repetitive (e.g. machine interactions, energy usage) and non-repetitive data (e.g. emails, warranty claims). Data is repetitive if records do not significantly differ from each other. For example, a sensor that measures the temperature of a machine every second, will most likely produce repetitive data, since under normal circumstances the temperature will stay constant. However, if an error occurs the temperature might rise or fall indicating that something is wrong. Non-repetitive data, is data were each record is unique. When analyzing emails, for example, it is likely that one differs from another. [13] The opportunity for business relevancy in repetitive data mainly comes from the detection of outliers (e.g. manufacturing control information that exceeds a threshold). [13]

Prepare Data (Transform)

Once relevant data is identified, it can be stored in different databases. Big Data poses a serious challenge to existing database technologies. Even if relational databases can capture the data, it is hard to get data out again to perform analytics. [14]

A literature review showed that multiple solutions for the storage of Big Data exist. Each type has its advantages and thus a solution must be chosen with great care and in consensus to the actual use case. In this research project a MySQL database for data storing is chosen. This concept has some limitations but it is the simplest and most cost efficient solution for this feasibility study.

Even though there are NoSQL solutions that provide consistency, there is another database type that might be better suited for use cases, where scalability and consistency are important. NewSQL is a promising new technology that tries to merge the best of the SQL and NoSQL world, but it does not come without trade-offs as well. However, the technology is promising to solve certain problems where time is of the greatest importance (e.g. real-time analytics). Again, different types of in-memory approaches lead to the fact that the use case must be taken into consideration when choosing a database.

Perform valuable analytics (Analyze)

Big Data also poses a challenge on the analytical side as more data can be used to perform different, more advanced types of analytics. Analytical tasks are used to improve the decision-making process [15] and thus represent the purpose of all the previous steps Big Data is involved with. Furthermore, a huge quantity of analytical tasks exist that are attributed to solving different problems and help with decision-making.

This research project is structured based on the theoretical approach described. The following explains the practical learning and investigation results achieved over the past few months.

3. Sensor Integration in the Environment of FabLab Graz

Based on our own experience the basic function of a FabLab is to provide rapid prototyping functionality. The research team thus decided in a first step to mount sensors inside the 3D printer, CNC mill, laser cutter and the FabLab room. These machines have a very high utilization and are the main tools for rapid prototyping. The main goal for the technical implementation is to design a cheap to maintain system, which is built from open source software and easily accessible hardware and which is also as extendable as possible. A first thought of the sensor system, which is implemented in the FabLab Graz, is shown in Figure 2.

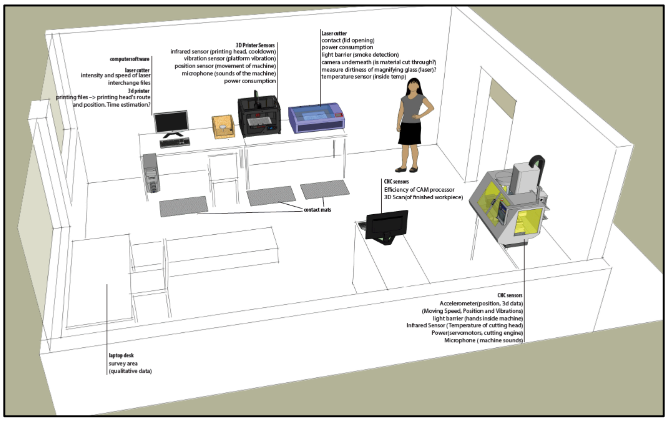

Figure 2: Sensor system in the FabLab Graz

As a platform, Arduino is selected. Ardunio is the most widely used physical computing platform. It offers digital and analog inputs as well as an I2C bus. There are many variations available, which differ in physical size, number of inputs and features. For this research project an Arduino Mega2560 and Arduino Uno is used.

Arduinos cannot store the quantities of data we are expecting internally. They are also limited to the aforementioned sensor inputs, but for the data gathering it is necessary to have the option to use microphones and webcams too. Thus, the Arduino is connected to a RaspberryPi, a small computer running Linux. Raspbian (raspbian.org) runs on RaspberryPi B+ and a 32 GB microSD card is used to have the option of storing measurements locally.

Firmata was selected for the communication between Arduino and RaspberryPi. Firmata is a protocol for communicating with microcontrollers from software on a computer. The Arduino software package already includes a routine (standard_firmata), which installs the protocol on the Arduino. The Python module (PyMata, github.com/MrYsLab/PyMata) is then used to communicate with the Arduinos on the RaspberryPis.

This setup not only allows the use of Python and avoidance of writing low level code in processing to run on the Arduino, but it also enables testing and debugging of all generated code by connecting an Arduino directly to computers. This represents an enormous advance over having to SSH into a RaspberryPi and debug from there. The same Python script runs on all RaspberryPis. To configure the sensor connection a .json-file is used. The implementation of the sensor specific software is discussed in the next section.

The first tests showed that the internal clocks of the RaspberryPis were not consistent enough and after a few hours without power, they set their time back to the start of the unix epoch. Thus a means of having a consistent time signal to all RaspberryPis is needed or any analysis involving sensors connected to two separate ones would have been meaningless.

The problem is solved by connecting all RaspberryPis to a simple network infrastructure. The network consist of RaspberryPis,

a Synology DS213 network attached storage and a switch. The time is set on the NAS, because it is the central hub of this system and also retains time the longest when it was without power. A MariaDB SQL Server runs on it. All RaspberryPis insert their data after each measurement with a query, attaching the server's timestamp to each row.

Discussion and learning

In general, the system worked well and was easy to maintain. A few problems occur, mostly because of the unsynchronized times and the lack of capability for sending log messages to the outside world, e.g. via e-mail to warn before a system runs out of memory. It is thus strongly recommended that an Internet connection for future projects of this type should be used.

Arduino has recently released the Arduino Yún, which resembles an Arduino Uno, but brings USB, Ethernet and Wi-Fi support. [16] This allows removing the RaspberryPis from the system and with Wi-Fi also avoids having network cables all over the lab. The use of Arduino Yún instead of the setup mentioned above will reduce the costs and the complexity of the system.

The most significant insight was that the I2C sensors do not work with the PyMata setup. Most of them worked with the processing libraries included, but their use did not dovetail with the declared system architecture.

4. Collecting, Analyzing and Processing of Data

Data from the three machines (3D printer, CNC mill, laser cutter) and the room environment was gathered over a period of two months. The data from all the sensors was recorded at approximately 1Hz. This is the maximum the Arduino can handle, because in the given setup it has to read one sensor and then store it before going on to the next sensor. This process for up to 15 sensors per machine takes up to 600ms. Additionally, a safety buffer is necessary. During the project, the team figured out that the selected sampling rate is far too high for some measurements that change very slowly such as temperature and far too low for measuring movements by integrating accelerator measurements.

In addition, the noise level of the raw data is too high for automatic pattern detection algorithms and makes it harder to see patterns visually. Thus the noise is reduced before the data analysis phase. Simple and exponential moving averages are thus used and their window size is adjusted to a visually pleasing result. Methods such as the Kalman filter would be a better choice or more sensors and a longer data recording period, because they self-adjust their smoothing intensity while we set it to a constant value for each sensor and chose this value manually. It is important to consider that all kinds of smoothing represent some types of weighted average, introducing a time lag into the system that depends on the window size. An exemplary result of the smoothing applied to the raw data is shown in Figure 3.

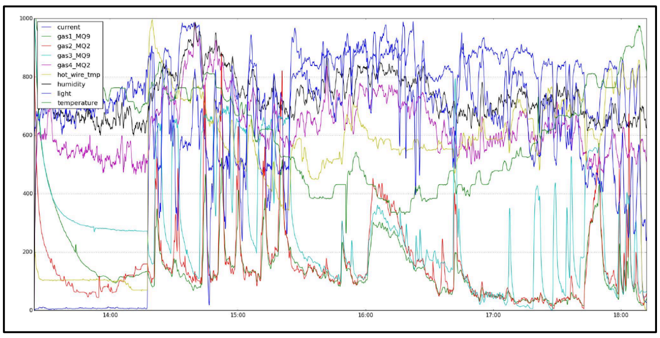

Figure 3: Data graph of laser cutter (5 hours)

R (r-project.org) and Python (python.org) are used for all processing, explorative analysis, model building and charting for the collected data in a tabular format from the SQL server. The screenshots of the CNC mill control software are transformed to text with tesseract-ocr (code.google.com/p/tesseract-ocr/).

Discussion and learning

The pyAudioAnalysis library (github.com/tyiannak/ pyAudioAnalysis) is used for the analysis of the recorded sound files. This provides methods for feature extraction and classification and is thus able to extract characteristics such as the sound intensity at a given time. The goal was to use this information as additional sensors. However, the time synchronization between the NAS and the RaspberryPis proved to be faulty and it was not possible to correct this in a short run. Nevertheless, the library's classification algorithm was successfully tried out on a small sample of 3D printer audio files.

An R and Shiny (shiny.rstudio.com) based plotting tool for the data analysis has been developed to allow every member of the research team to take part in the analysis (see Figure 4).

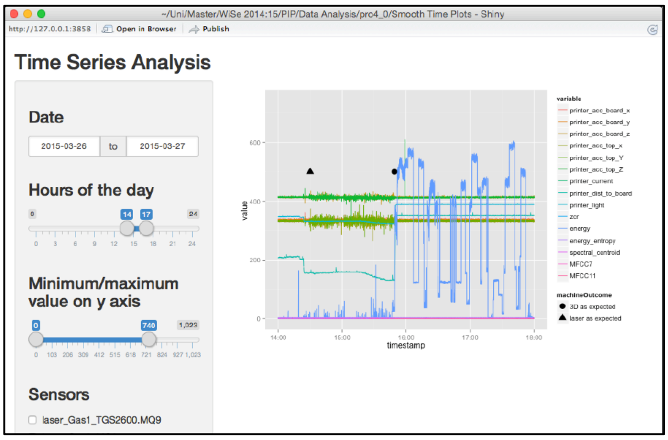

Figure 4: Coded plotting tool based on R and Shiny

R and Python are the standard tools for the required tasks and are capable of everything that is needed. Python has the advantage of being faster, while R is more interactive. The software IPython (ipython.org) and IPython notebook (ipython.org/notebook) are a good start for learning about interactive Python. It is possible to perform standard tasks for optical character recognition (OCR) and sound, but optimization of the sensor setup was necessary to run more advanced analysis (e.g. transforming the width of a progress bar to a value classify machine movements by characteristic sounds).

5. Derivation of Use Cases and Product Improvements Based on Data Correlation

This project phase is characterized by discovery correlations and patterns in the data obtained from machines and the FabLab room environment. The R programming language is used as a base of operation. Specific data phonemes were discussed in various workshops. In addition, interesting data correlations and potentials for product enhancements were worked out in the course of interviews with the FabLab staff as also with users.

Predicting service intervals and emergency stop

The data analysis showed that the activity of the laser cutter (measured by its electric current) is immediately followed by a deterioration of the air quality at the laser cutter and its ventilation system, and after continuous usage a slight deterioration of air quality in the room (both measured by the various gas concentration sensors).

When analyzing this phenomenon over all recorded days, however, it is noticeable that the negative effect on the room air

quality worsens over time and then suddenly disappears. At the end of the recorded time period, the effect of the laser cutter usage on the air quality in the room mostly fades (see Figure 5).

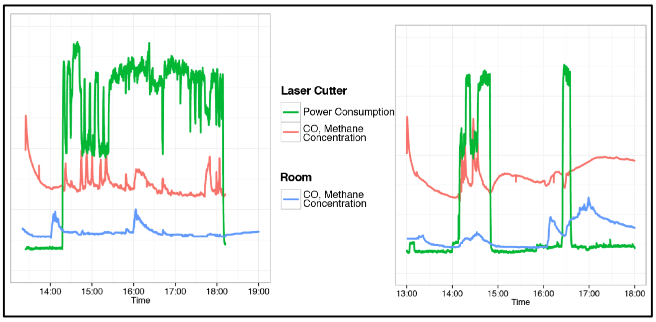

Figure 5: Laser cutter before filter change (left) and after (right)

It turned out that the filter in the ventilation system was changed on the exact day the phenomenon in the data disappeared. This filter needs to be changed regularly. Changing it too late leads to health concerns for the FabLab staff and users as well as a decrease in the cutting quality. Combustion gases can accumulate in the cutter chamber and hinder visual inspection of the cutting process while also posing a danger to the focus lens, which can become unusable if smoke particles burn into the lens. Cleaning it too early creates unnecessary waste, by contrast cleaning too late results in a broken and unusable lens, which is a costly negative development.

Good working conditions and an unconsumed filter mean that the room air quality should not decrease when the laser cutter is running and if the filter is used over a longer period it decreases. A simple use case is derived from this finding. It predicts a filter change before it is really required and notifies the user in good time to order replacement parts, in the same manner as a business printer prompt. Furthermore, with the developed setup it is possible to perform an emergency stop of the laser cutter automatically, if the ventilation system has a defect or if the cut material develops combustion flames.

User notifications and cost reductions

The CNC mill runs tasks that often take several hours and are started mostly late in the afternoon (in the case of FabLab Graz). The manufactured part is inspected at the earliest on the following morning or sometimes even one day later. Data analysis shows that the current used by the machine is only negligibly smaller when the machine is turned on but not running compared to when it is running. Added to this the CNC mill needs compressed air for its tool change mechanism. It leaks this compressed air, whenever the machine is turned on or off.

Based on this finding an algorithm that recognizes when the machine operation has finished is coded. The sound patterns of the machine and its power consumption are analyzed for this purpose. This use case can further be improved by having more sensors installed, or by including real time OCR of the machine control software. Based on this it is possible to have a robust detection of production finish and automatic cut-off for the power and compressed air. The power can be handled with a simple electric relay and an electrical valve can be used for the compressed air.

Displaying and predicting FabLab usage

It was not possible during the project period to record a sufficient amount of data to make statistically sound statements, but an interesting observation in this context is that more people attend FabLab Graz later in the afternoon and during the early evening. This makes sense, as students usually do not have classes at these late afternoon and evening times. Using the proposed sensor system architecture a prediction can be made of the number of persons actually present in the FabLab. Furthermore machine downtimes can be recorded. For instance, this information can be displayed on the FabLab website to help users to find blank spots of machine usage or to inform them if a machine is currently occupied. As a result, this use case can help to increase FabLab utilization and user satisfaction.

6. Conclusion and Outlook

The research team has managed to conduct an experiment and implement a Big Data analysis in the FabLab Graz and proved that this concept is not only feasible but also generates a very significant benefit for FabLab operators and users. In addition, the strong interest shown by many industrial companies demonstrate that the potential of the project is also significant and valuable. Based on these results this project will be continued.

This research visualizes the potential in developing a plug-and-play sensor kit with standardized open source software and hardware in order that every FabLab can benefit from the machine data gathered. Therefore, a first prototype of such a system was developed during the project. (see Figure 6) The next step is that the developed plug-and-play sensor kit not only gathers machine data in the FabLab, but also stores the data automatically on a webserver to provide this data for the community.

Figure 6: First prototype of a plug-and-play sensor kit

The goal is to implement the developed senor kit in many other FabLabs worldwide in order to collect and provide a very large quantity of data for use in further research work. In the spirit of open data, everybody will be able to access the database. This will make a significant contribution to the FabLab community. The three use cases referred to give a first insight into what is possible even with a very limited amount of data. It can only be imagined what the possibilities will be with the availability of a bigger data set. Hopefully, more fascinating correlations may be found in the near future.

Acknowledgement

This research has been partly carried out by a student team of the course “Product Innovation Project” at the Graz University of Technology hosted by the Institute of Industrial Management and Innovation Research.

7. References

[1] fablabs.io, [Online]. Available: https://www.fablabs.io/labs. [access date 20 7 2015].

[2] fab.cba.mit.edu, [Online]. Available: http://fab.cba.mit.edu/about/faq/ [access date 20 7 2015].

[3] Rania, M., & Mahmoud, A., Emphasizing the advantage of 3d printing technology in packaging design. International Design Journal , 2012, 1, p.111-119.

[4] Big Data Analytics, Ericsson White Paper, [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-big-data.pdf. [access date 20 7 2015].

[5] Communication Promoters Group of the Industry-Science, Recommendations for implementing the strategic initiative Industrie 4.0. Office of the Industry-Science Research Alliance., 2013.

[6] Mayer-Schönberger , V., & Cukier, K., Big data, a Revolution That Will Transform How We Live, Work, and Think, Eamon Dolan, 2013.

[7] Koch, V., Geissbauer, R., Kuge, S., & Schrauf, S., Industry 4.0 Opportunities and Challengies of the Industrial Internet, Strategy&PwC, 2014.

[8] Krishnan K., Data Warehousing in the Age of Big Data, Morgan Kaufmann, 2013.

[9] Kanetkar S., Chanchlani N., Artificial Intelligence and Cognitive Analytics Approaches towards Efficient Predictions for Business Intelligence, International Journal of Computer Applications, 103(14), 2014, p. 35-39.

[10] Minelli M., Chambers M., Dhiraj A., Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, John Wiley & Sons, 2013.

[11] Rockwelll Automation, Microsoft, [Online]. Available: http://customers.microsoft.com/ CustomerStory.aspx?recid=19922. [access date 20 7 2015].

[12] Marín-Ortega P.M., ELTA: New Approach in Designing Business Intelligence Solutions in Era of Big Data, Procedia Technology, 2014, p. 667-674.

[13] Inmon W.H., Linstedt D., Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault, Morgan Kaufmann, Burlington., 2014.

[14] Jacobs A., The Pathologies of Big Data, in: Communications of the ACM, 52(8), 2009, p. 36-44..

[15] Al-ma M.A., The Role of Business Intelligence Tools in Decision Making Process, International Journal of Computer Applications, 73(13), 2013, p. 24-31.

[16] Arduino Yún, Arduino, [Online]. Available: https://www.arduino.cc/en/Main/ArduinoBoardYun? from=Products.ArduinoYUN. [access date 20 7 2015].